DOCUMENT RESUME

ED 088 614 PS 007 241

AUTHOR Montague, James C., Jr.
TITLE Preliminary Computer Verbal Content Analysis in Preschool Black Children.
PUB DATE [73]
NOTE 22p.

EDRS PRICE MF-$0.75 HC-$1.50
DESCRIPTORS Age Differences; *Computational Linguistics; Computer Science; *Content Analysis; Language Development; *Negro Youth; *Preschool Children; *Speech

ABSTRACT

The major purpose of this study was to investigate expressive content language differences in preschool black children that might relate to Piaget's theory of egocentric speech. Correlating first order Harvard III Psycho-sociological Dictionary categories with Piaget's egocentric speech and the second order categories with socialized speech it was hypothesized that few or no second order content categories would differ within contrasted groups of 3-, 4-, and 5-year-old black children. A measure of structural syntactical change was also made for comparison purposes. Statistical test results confirmed this hypothesis. Computerized verbal content analysis appears to be a fruitful method for testing and studying theoretical issues and empirical research in child development. (Author/DP)

ED 088614

JAMES C. MONTAGUE, JR.

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PRELIMINARY COMPUTER VERBAL CONTENT ANALYSIS IN PRESCHOOL BLACK CHILDREN

James C. Montague, Jr.

University of Arkansas at Little Rock

PS 007241

Address Correspondence to:

James C. Montague, Jr.

Director, Communicative Disorders

University of Arkansas at Little Rock, Little Rock, AR 72204

ABSTRACT

The contentive verbal language of groups of three-, four-, and five-year old Black children was contrasted by utilizing the computerized General Inquirer System and the associated Harvard III Psycho-sociological Dictionary. Correlating denotative Harvard III categories with Piaget's egocentric speech and connotative categories with socialized speech it was hypothesized that few or no contentive categories would differ between the three age groups of children. Statistical test results confirmed this hypothesis.

Computerized verbal content analysis appears to be a fruitful method for testing and studying theoretical issues and empirical research in children.

JAMES C. MONTAGUE, JR.

It would appear possible that changes in childrens' developmental cognitive styles could be reflected through their expressive verbal language. Innumerable studies have been made of the structural, descriptive linguistic elements, e.g. phonology, morphology and syntax, in childrens' language acquisition patterns. These structural language studies have provided us with important information in child development; for example articulation development in normal children and structural language differences and omissions in retarded groups of children. However, these studies for the most part have not been correlated with chronological cognitive development.

One notable exception to this are the studies that Piaget and his colleagues have conducted in ego-centrism among young children. Piaget (1959) divides child language into roughly two large referent groups -- the ego-centric and the socialized.

> When a child utters phrases belonging to the first group, he does not bother to know to whom he is speaking nor whether he is being listened to. He talks either for himself or for the pleasure of associating anyone who happens to be there with the activity of the moment. This talk is ego-centric, partly because the child speaks only

about himself, but chiefly because he does not attempt to place himself at the point of view of his hearer (pp 9).

Ego-centric speech is further divided into the three categories of repetition (echolalia), monologue, and dual or collective monologue. In ego-centric speech the young child does not take into consideration the point of view of the other person; his presence serves only as a stimulus. On the other hand socialized speech does take into consideration the point of view of the other person and contains adapted information, criticism, commands, requests, threats, questions and answers. A proportion of ego-centric speech is obtained by dividing a child's number of ego-centric remarks by other spontaneous forms of language. This is reported in coefficients of ego-centrism which Piaget's colleague Mme Leuzinger reported for the following age groups of children:

AGES	AVERAGE COEFFICIENTS OF EGO-CENTRISM
3 years	0.51
4 years	0.48
5 years	0.46
6 years	0.45
7 years	0.28

Thus, as shown above Mme Leuzinger has indicated that ego-centric speech, between the ages of three to six, passes through a semi-stationary phase during which it gradually decreases while at the same time it

fluctuates between one-half and one-third of the total amount of speech; after the age of 7, ego-centric speech finally tends to decrease to less than one-quarter of the totality of spontaneous speech.

One of the difficulties with applying Piaget's language system relates to manual processing time and reliability problems that prohibit observations of any but small quantities of verbal data. Other verbal content analysis systems, like Piaget's, have likewise been restricted in application -- primarily because manual coding has proved to be tedious, monotonous, time-consuming, expensive and unreliable. In 1961, Stone and Bales developed the initial version of the General Inquirer computer content analysis system. Words are pigeon-holed or categorized into specific content categories. The labels over the pigeon-holes -- that is, the categories referred to in different content dictionaries -- can be arbitrarily changed without impairing the overall efficiency of the General Inquirer system. The proper selection or construction of the content dictionary becomes the most crucial aspect of content analysis. At this point, the raw verbal data is tied to theory and a basis for drawing conclusions and making inferences is achieved.

The computer content analysis dictionary that appears to be most compatable for the study of Piaget's ego-centric language theory in children is the Harvard III Psychosociological Dictionary. This is a general purpose content analysis dictionary that can be used in conjunction with research of a psychological and sociological character. In addition to its proven application with adults, this dictionary has

been successfully used for studying contentive aspects of language in middle-class white children (Pine, 1970) and in comparing institutionalized versus non-institutionalized retarded children (Montague, et al., 1974). In constructing this dictionary, Dumphy (1966) stated:

> . . .sociology furnished a set of categories better suited to classifying roles, objects, and cultural artifacts than did psychology. On the other hand, psychology presented more clearly defined categories for dynamic processes. We proceeded, therefore, to give most nouns (object names) a sociological definition and most verbs a psychological definition at the denotative level.

In collating Piaget's egocentric speech theory with the Harvard III computer content system we turn to another division of the dictionary into denotative andconnotative levels. Referral to Table 1 indicates that the content categories are divided into first-order and second-order tags or categories. First order tags represent the most explicit primary denotative meaning of words. Second-order tags fill out the limited meaning given by the assignment of a single first-order tag. In this dictionary second-order tags are designed to identify generalized concern with the external or internal worlds. Thus second-order tags, or content categories, correlate closely with Piaget's more mature socialized language that the developing child acquires in a greater proportion as he/she develops in chronological age.

(Approximate position of Table 1)

As in any unique linguistic application to preschool children, the tradition has been to study only the verbal behavior of middle-class, white children. Currently there is a paucity of normative data on language variables in minority children. What studies that have been conducted with Black children generally do not delve beneath the age of preadolescents or if they do study preschool Black children are concerned with structural language elements e.g. Henrie, 1969.

The major purpose of this study was to investigate expressive content language differences, as measured by the Harvard III Dictionary, in preschool Black children that might relate to Piaget's theory of egocentric speech. Referring back to Leuzinger's data on three, four, and five-year olds we find relatively high and stable proportions of ego-centric language patterns. Correlating first order Harvard III categories with Piaget's egocentric speech and the second order categories with socialized speech it is hypothesized that few or no second order content categories will differ within contrasted groups of three, four and five year old Black children. Likewise, there should be few or no differences between these age different Blacks in first-order content categories representing egocentric speech.

While the major purpose of this study was to look at contentive changes in language acquisition among different age groups of preschool Blacks it was also felt that some measure of structural syntactical change should be made for comparison purposes. While several syntactical language analysis systems are available a system was selected that would appear to reflect chronological, developmental age linguistic changes.

PS 002741

The methodology utilized in this study was a modification of Lee's (1966) developmental sentence type system. Lee presents four major types of sentence patterns: two word combinations, noun phrases, constructions and kernal sentences. These are presented in an emergin, hegemony language acquisition type of system among young children. For this portion of the study two additional categories; single word utterances and unclassified were added.

METHOD

Subjects

The subjects consisted of three, four and five year old Black children enrolled in preschool urban and rural daycare programs in Alachua County, Florida. These subjects were selected on a trimodal, closely related, chronological age basis. This data is contained in Table 2.

(Approximate position of Table 2)

While IQ data were available this was not used for matching the groups due to the white culture bound nature of the IQ test scores available. None of the subjects displayed any gross sensory, perceptual or motor deviations.

Verbal Language Gathering Procedures

It was recognized that an unstructured "slice of life" language gathering technique using portable body harness microphone/transmitters, while the children engaged in free play, might be the most valid method of sampling verbal language from these children. However, pilot testing, revealed that competing childrens' voices and ambient noise created a "Watergate tape effect" with serious transcription reliability problems negating any possible validity advantages.

In an effort to deal with this problem a semi-structured "tell-a-story" procedure evolved. Stimulus materials consisted of a total of 20 specially prepared 2" by 2" colored slides. Thirteen of these slides depicted Black preschool children in various play activities. The remaining seven slides were reproductions of language pictures from the Peabody Kit (Level 1). These stimulus materials proved to be quite effective in stimulating a sufficient verbal language sample from each child.

An automatic Kodak carousel slide projector, with a wide angle lens and remote control switch, was used for presenting the stimulus materials. A Sony tape recorder was used for recording each child's responses to the stimulus slides. Scotch brand .5 mil magnetic tape was used for recording purposes at a speed of 3 3/4 inches per second.

The recorder and slide projector were set up in a quiet, partially darkened room, in each daycare center, with a white wall functioning as a satisfactory screen. Efforts were made to gather the recordings at those periods of the day when outside ambient noise was at a minimum. Two examiners were trained in a standard procedure for gathering the data. The child being examined was brought into the room and seated in a child size chair. The examiner, seated next to the child, said, "We are going to look at pictures and tell stories about them. The first slide was then projected on the wall. The majority of subjects would begin to verbalize with only minimum verbal reinforcement being used by the examiner. For a few children additional encouragement in the form of "What do you see?, What is happening?" was required.

Transcribing Procedures

The total number of words collected by subject and group is listed

in Table 2. The recorded verbal language samples were transcribed into ordinary English orthography by a specially trained secretary. A special Wollensak Model 1600 tape recorder, with a foot pedal back up switch, was used by the transcriber in conjunction with an electric typewriter for the majority of the transcriptions. All verbal remarks, of both the examiner and individual subject being recorded, were then typed out. When an unclear word was encountered an underlined blank was left in its place. Through random retyping of previously typed transcripts it was determined that there was good intra-transcriber reliability. It was also established that inter-judge reliability was good by having two individuals type up transcripts and compare them with samples from the specially trained transcriber.

Data Coding and Processing

After completion of the typed transcriptions the investigator functioned as the syntax coder classifying each utterance of all the subjects with a numerical subscript. These subscripts were coded to signify if the utterances were single-word, two-word combinations, noun phrases, constructions, kernal sentences or unclassified. Some additional minor conventions in editing were also necessary in order to prepare the typescriptsfor the computer content analysis.

Tagging Additional Words

It was recognized that many of the more frequent words of Black preschool children might not be contained in the Harvard III Psychosocial Dictionary. A random sample of several Black preschool language samples was processed through a Frequency Word Listing Program developed by Hutchinson and Lynch (1970). This provided an alphabetic printout,

by frequency, of all the words contained in the samples. The printout from this program was then compared with a printout of the Harvard Dictionary. Those high frequency words from the preschool Black childrens' language samples not occuring in the dictionary were then appropriatdy coded, key punched and inserted into the Harvard Dictionary. Table 3 contains a listing of the additional high frequency words added to the dictionary.

(Approximate position of Table 3)

In assigning these non-dictionary, high frequency words to specific content categories, reference was made to similar Harvard III words already coded. For example "truck" was coded the same as "Automobile."

Computer Content Analysis Processing

The verbal data samples were key punched on standard 80 column electronic data processing cards which were verified for reliability. At this point the data were processed, individually by subject, through the General Inquirer and content category totals obtained. This computerized procedure consisted of the number of times the texts' content words were classified into each of the Harvard Dictionary categories. The content category tally operation resulted in the assignment of two types of numbers to each document analyzed; one set represented the absolute frequency of occurrence (raw score); the second represented the relative frequency of occurence (index score) of the content categories.

RESULTS

For analysis, the word index content category scores were used for comparison. These index scores offer the advantage of not being influenced by the length of the verbal corpus, such as indicated by the variation

between the various age groups in this study. Due to the large number of zero scores, in certain content categories, nonparametric statistical techniques were used in this analysis. The Kruskal-Wallis one-way analysis of variance by ranks (Kirk, 1969, pp 493-94) provided an overall test of significance between the three separate age groups of the preschool Black children. In order to determine which comparisons among the three treatment groups were significant Dunn's (1964) process of multiple comparisons using the rank sums procedure was employed. Table 4 presents the significant syntactical and contentive categories favoring specific age groups of the Black children.

(Approximate position of Table 4)

Under the syntactical categories Developmental Sentence Type 1, which represents single word responses was significantly more prevalent in the verbal language of the three-year old group as compared to the four-year olds. Table 4 also indicates that Kernal Sentences are much more frequent in the language patterns of five-year olds in contrast to three-year olds.

Within the contentive categories both the three and four-year old preschool Blacks had significantly more Male-Role words in their expressive language than did the five-year old group. In contrast the five-year group had scores significantly higher on Message Form and Expel content words in comparison to the three-year old group. It was also observed that the five-year olds also had more Community content words than the four-year old group. A final contentive difference was observed in the Family category with the three-year old group scoring higher than the four-year old group.

DISCUSSION

The results confirm the hypothesis that application of first and second order Harvard III contentive categories, respectively correlating with Piaget's egocentric and socialized speech, largely results in no differences between contrasted groups of three, four and five-year old Black children. A total of six syntactical categories and sixtynine contentive categories were analyzed between the three different age groups of preschool Black children. Considering the extensive number of orthogonal comparisons tested it becomes evident that the eight differences reported in Table 4 is not large. While these differences might represent Type II statistical differences it might be wise to make some evaluative comments.

The Male-Role content category was certainly more prominant in the younger three- and four-year old children than in the older five-year old children. Contentive words, used by these children contained in this category, included "boy, man, he, daddy." It should be noted that a majority of the stimulus pictures were of Black boys engaged in various types of play activities. It appears that the frequency of the word "boy," within the corresponding Male-Role category, decreases as the child increases in age. Examination of the language samples indicates that there is less noun naming in the stimulus pictures, i.e. "There a boy" or "boy" and more action i.e. "Be playin wif truck" as the children ascend into their fifth year.

Other content categories showing significant differences included Message-Form and Expel which favored the five-year olds in contrast to the three-year olds. Message-Form relates to names of communication

media, in a broad sense including nouns referring to types of objects involved in communication. Words used among the five-year old Blacks in this category included "name, book, case, point, story, and sign." Quite possibly the five-year old has significantly more outer-directiveness in his cognitive development which displays itself verbally with more Message-Form words.

In addition the five-year olds had significantly more words in the Expel category than the three-year olds. Expel category words used more frequently by the five-year olds included "push, throw, and drop" and represents the process of ejecting and is the final category of an instrumental action theory. Possibly this represents more action type identification and activity emerging in the five-year olds.

The five-year olds were also superior, in comparison to the four-year olds, in the Community category. This is an institutional content category and inspection indicates that the five-year olds used more community based words such as "people, park, name and hello." This is a second-order Harvard III Dictionary category on the connotative level. With particular reference to the word "people" this category, like the previously mentioned Message-Form category, might represent a developing "outer direction" of socialized language in the maturing five-year old Black child that is not present in younger children.

A final contentive category Family was more prevalent in the language of the three-year old group in contrast to the four-year olds. This category contains words like "mother, daddy, home, chair and children" that directly relate to the family environment. By far the major word used in this category was "mother" and its derivatives.

Apparently between the ages of three and four -- preschool Black children begin to deal more abstractly with the stimulus pictures; for example referring to a woman as a "woman" rather then as a "Mama."

Turning to the syntactical comparisons the single word responses were more frequent in the language patterns of the three-year old children in contrast to the four-year old group. This represents more single word naming on the part of the three-year old children. The kernal sentences, as postulated under Lee's system, were more prevalent in the language samples of the five-year old Blacks in contrast to the younger three-year olds. This acquisition of higher level kernal sentence forms by preschool Blacks at age five would follow a naturally developing performance level of syntactical development.

ACKNOWLEDGMENT

Special assistance was rendered by Miss Karen Silvernail in data collection and Mrs. Betty Smith was the patient data transcriber. The author would like to thank Dr. Joann Strickland and Dr. Edward C. Hutchinson for their support during this project. A special note of thanks is due Mrs. Mary Lynch, former Computer Programmer, University of Florida Computer Center, for her help with the various computer language programs used in this study. The author was supported during this research by an Office of Education Fellowship.

REFERENCES

Dunn, O. J., Multiple comparisons using rank sum. Technometrics, 1964, 6, 241-252.

Dunphy, C., The construction of categories for content analysis dictionaries, in P. J. Stone (Ed.) The General Inquirer: A Computer Approach to Content Analysis. Cambridge: The M.I.T. Press (1966)

Henrie, S. N., Jr., A study of verb phrases used by five year old non-standard Nego English - speaking children. Unpublished doctoral dissertation, University of California at Berkeley, 1969.

Hutchinson, E. C., and Lynch, M. C., Frequency word listing program, computer program developed at the University of Florida Computer Center, Gainesville, Fla. (1970).

Kirk, R. E., Experimental Design: Procedures for the Behavioral Sciences. Belmont, Calif.: Wadsworth Publishing Co., Inc. (1969).

Lee, L. L., Developmental sentence types: a method for comparing normal and deviant syntactical development, J. Speech & Hearing Dis., 32. Nov. 68, pp 311-330.

Montague, J. C., Hutchinson, E. C. and Matson, E., Computerized verbal content analysis of institutionalized versus community retarded children. (in press) J. Speech and Hearing Research, 1974

Piaget, J., The Language and Thought of the Child, The Humanities Press, 1959.

Pine, S., An analysis of the content of oral language patterns of children. Ph.D. Dissertation, University of Florida, Gainesville, Fla. 1970.

Stone, P. J., Dumphy, D. C., Smith, M. S. and Ogilvie, D. M., The General Inquirer: A Computer Approach to Content Analysis: Cambridge: The M.I.T. Press, 1966.

Table 1. Harvard third psychosociological dictionary content categories.

FIRST-ORDER TAGS

Objects

Social Realm

-Persons

Self
Selves
Other

-Roles

Male-Role
Female-Role
Neuter-Role
Job-Role

-Collectivities

Small-Group
Large-Group

Cultural Realm

-Cultural Objects

Food
Tools
Clothing

-Cultural Patterns

Ideal-Value
Deviation-Values
Action-Norm (norms)
Message-Form
Thought-Form (concepts)
Nonspecific-Object

Cultural Realm (cont.)

-Cultural Setting

Social Place

Natural Realm

Body Part
Natural Object
Natural World

Qualifiers

Time Reference
Space Reference
Quantity Reference
Quality (Sensory) Reference

Processes

Psychological Processes

-Emotions

Arousal	Anger
Urge	Pleasure
Affection	Distress

-Thought

Sense	If	Equal
Think	Not	Cause

Processes (cont.)

Psychol. Processes (cont.)

-Evaluation

Good	Ought
Bad	

Behavioral Processes

-Social-Emotional Actions

Approach	Avoid
Guide	Follow
Control	Communicate
Attack	

-Instrumental Actions

Attempt	Get
Work	Possess
Move	Expel

SECOND-ORDER TAGS

Institutional Contexts

Academic
Artistic
Community
Economic
Family
Legal
Medical
Military
Political
Recreational
Religious
Technological

Status Connotations

Higher-Status
Peer-Status
Lower-Status

Psychological Themes

a. Overstate
   Understate
b. Sign-Strong
   Sign-Weak
c. Sign-Accept
   Sign-Reject
d. Male-Theme
   Female-Theme
e. Ascend-Theme
f. Authority-Theme
g. Danger-Theme
   Death-Theme

Table 2. Sex, chronological age, and number of words collected by subject for the three age groups.

Sex	3 year olds (N = 12)	Number of Words Collected	Sex	4 year olds (N = 12)	Number of Words Collected	Sex	5 year olds (N = 12)	Number of Words Collected
M	3-6	151	F	4-3	302	M	4-10	307
F	3-6	184	M	4-3	230	F	4-10	501
F	3-6	434	F	4-4	307	F	4-11	476
M	3-8	107	M	4-4	141	M	4-11	347
F	3-8	595	F	4-5	355	F	5-0	417
M	3-8	287	M	4-5	196	F	5-1	465
M	3-9	357	F	4-6	490	M	5-1	156
M	3-9	134	F	4-6	117	F	5-2	343
M	3-10	417	F	4-6	365	M	5-2	367
F	3-11	137	F	4-7	460	M	5-3	447
F	4-0	470	M	4-8	384	F	5-3	217
M	4-0	183	M	4-8	486	F	5-1	514
Range	3-6 to 4-0	107 to 595		4-3 to 4-8	141 to 490		4-10 to 5-3	156 to 514
Mean	3-9	288		4-5	319		5-1	380

Table 3. High frequency preschool Black children's words inserted into the Harvard dictionary.

Word	First-Order Tag	Second-Order Tag
truck	Tool	Male-Theme, Economic
bus	Tool	Male-Theme, Economic
momma	Female-Role	Family, Higher-Status
motorcycle	Tool	Male-Theme, Recreational
alligator	Natural-Object	Sign-Strong, Danger-Theme
rabbit	Natural-Object	
airplane	Tool	Male-Theme, Economic
clown	Male-Role	Recreational
monkey	Natural-Object	
suitcase	Tool	
sandbox	Social-Place	Recreational
bicycle	Tool	Recreational
bike	Tool	Recreational
tractor	Tool	Male-Theme, Economic
jumprope		Recreational
ain't	Not	
outdoors	Space-Reference	

Table 4. Summary of significant statistical tests for syntactical and contentive differences between 3 year old, 4 year old and 5 year old preschool Black children.

Significantly more in:	3's as compared to 4's	3's as compared to 5's	4's as compared to 5's	4's as compared to 3's	5's as compared to 4's	5's as compared to 3's
			Contentive Categories			
Male-Role		.05	.06			
Message-Form						.02
Expell						.03
Community					.05	
Family	.02					
			Syntactical Categories			
Developmental Sentence						
Type I (One Word)	.03					
Kernal Sentences						.03